DOI:
10.1103/PhysRevLett.113.238106

Extensive parallel processing on scale free networks: Supplementary Information

Peter Sollich,[1] Daniele Tantari,[2] Alessia Annibale,[3,4] and Adriano Barra[5]

[1]*Department of Mathematics, King's College London, Strand, London WC2R 2LS, U.K.*
[2]*Dipartimento di Matematica, Sapienza Università di Roma, P.le Aldo Moro 2, 00185, Roma, Italy.*
[3]*Institute for Mathematical and Molecular Biomedicine, King's College London, Hodgkin Building, London SE1 1UL, U.K.*
[4]*Mathematics Department, King's College London, The Strand WC2R 2LS, London, UK.*
[5]*Dipartimento di Fisica, Sapienza Università di Roma, P.le A. Moro 2, 00185, Roma, Italy.*

(Dated: November 10, 2014)

In this supplementary information we show details of how the scale-free networks are built within our approach, and we detail some of their properties including significant clustering. We provide background information on the validity of the cavity field technique used in our method, and finally give details of the analysis of the case of asymmetric patterns.

PACS numbers: 07.05.Mh,87.19.L-,05.20.-y

1. *Hebbian scale free networks*

Our analysis concerns bipartite networks of neurons $i = 1,\dots,N$ and patterns $\mu = 1,\dots,P = \alpha N$. The patterns define Hebbian interactions $J_{ij} = (1/N)\sum_{\mu=1}^{P}\xi_i^\mu\xi_j^\mu$. The aim of this section is primarily to provide additional details of the structure of the resulting neural network, where two neurons i and j are thought of as connected by an edge if $J_{ij} \neq 0$. This network, which we shall denote generically by $\mathcal{G}$, can be regarded as the result of a marginalization (over patterns) of the bipartite network.

To construct the bipartite network, we can in general assume that the distribution $P(d)$ of the neuron-to-pattern degrees and the distribution $P(e)$ of pattern-to-neuron degrees are given as constraints, and networks are sampled uniformly across the set of networks satisfying these constraints. Our analysis in the main text applies to this general scenario. Here we focus more specifically on constraining only the pattern degrees, by positing a pattern-dependent dilution of the links $P(\xi) \propto \prod_{i,\mu} P(\xi_i^\mu)\prod_\mu \delta_{e_\mu,\sum_i|\xi_i^\mu|}$ with

$$P(\xi_i^\mu) = \frac{e_\mu}{2N}(\delta_{\xi_i^\mu,1}+\delta_{\xi_i^\mu,-1}) + (1-\frac{e_\mu}{N})\delta_{\xi_i^\mu,0} \qquad (1)$$

This leads to a degree distribution $P(d) = \text{Poisson}(\alpha\langle e\rangle)$ for the neurons, while the degree distribution of the patterns is by construction $P(e) = P^{-1}\sum_\mu \delta_{e,e_\mu}$.

To generate such a network numerically, we first sample for each pattern a degree e_μ from the prescribed distribution $P(e)$, and then choose from a uniform distribution which entries of ξ_i^μ of the pattern are nonzero. The nonzero entries are chosen as $+1$ or -1 with equal probability as specified by (1). To obtain scale-free networks we choose specifically a power law degree distribution $P(e) \sim e^{-\gamma}$ (for $e \geq 2$) for some fixed $\gamma > 2$ [1]. To avoid unphysical pattern degrees we impose $e \leq N$, and accordingly normalize $P(e)$ so that $\sum_{e=2}^{N} P(e) = 1$.

Nontrivially, the neural networks generated – via Hebbian interactions – from such scale-free pattern degree distributions have high modularity (see Fig. 1, main text). We show below that they also have scale-free degree distributions (see Fig. 1) and exhibit significant clustering (see Fig. 2).

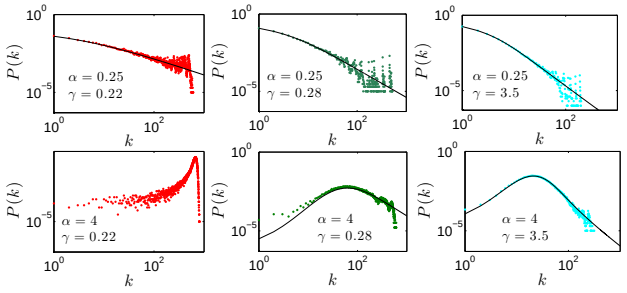

FIG. 1: Degree distributions $P(k)$ of neural networks $\mathcal{G}$ of size $N = 1,000$ as obtained for different values of the parameters α and γ. Each parameter combination is represented by a different color and symbol; see Fig. 2 for legend.

To understand the degree distribution in the neural network, one can first look at the limit $N \to \infty$, where this distribution can be predicted analytically. We assume as before that pattern μ has e_μ nonzero entries, with e_μ drawn from some distribution $P(e)$. Each pattern is then connected to a random subset of e_μ neurons. As each neuron i typically only receives connections to a finite number d of patterns μ in this way, overlaps between the connections created by the different μ are subleading for large N. One can therefore write for the number of neurons k that a given neuron interacts with, i.e. its degree within the neural network $\mathcal{G}$, $k = \sum_\mu c_\mu(e_\mu - 1)$ where $c_\mu \in \{0,1\}$ indicates whether pattern μ connects to the given fixed neuron and is $= 1$ with probability e_μ/N and $= 0$ with probability $1 - e_\mu/N$. The generat-

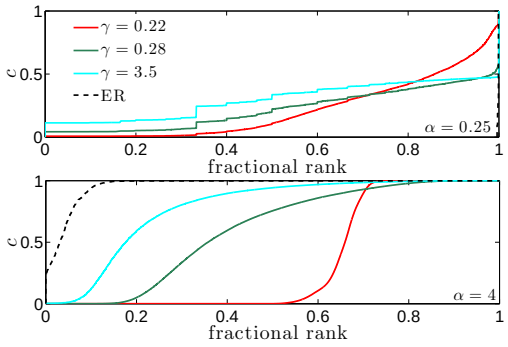

FIG. 2: Distributions of local clustering coefficients for neural networks of size $N = 1,000$. We plot clustering coefficients c_i against their (fractional) rank in a size-ordered list, i.e. effectively a cumulative histogram. Different symbols and colors represent different values of the parameters α and γ as shown in the legend. For the $\gamma = 3.5$ cases we also show the corresponding result for Erdös-Rény graphs with the same average neuron-to-neuron degree $\langle k\rangle$.

ing function of k [2] is then

$$\langle z^k\rangle = \prod_\mu \langle z^{c_\mu(e_\mu-1)}\rangle \quad (2)$$

$$= \exp\left(\sum_\mu \ln\left[1+(e_\mu/N)(z^{e_\mu-1}-1)\right]\right) \quad (3)$$

For large N one can expand the argument of the logarithm to get

$$\langle z^k\rangle = \exp\left(\alpha\,\frac{1}{P}\sum_\mu e_\mu(z^{e_\mu-1}-1)\rangle\right) \quad (4)$$

The average over μ approaches an average over $P(e)$ for $N\to\infty$, so that

$$\ln\langle z^k\rangle = \alpha\langle e(z^{e-1}-1)\rangle \quad (5)$$

To write the average on the r.h.s. more simply one can define $\tilde e = e-1$ and $\tilde P(\tilde e) = eP(e)/\langle e\rangle$:

$$\ln\langle z^k\rangle = \alpha\langle e\rangle\left(\langle z^{\tilde e}\rangle - 1\right) \quad (6)$$

For the Pareto case $P(e)\propto e^{-\gamma}$ for $e\geq 2$, one gets $\tilde P(\tilde e)\propto(\tilde e+1)^{1-\gamma}$ for $\tilde e\geq 1$. The intuitive interpretation is that for a randomly chosen neuron-to-pattern edge, $\tilde e$ is the number of *other* neurons the patterns is connected to. Its distribution is proportional to $eP(e)$ because patterns with e connections have e chances of being chosen randomly.

The large-k asymptotics of $P(k)$ can now be deduced without difficulty. In the average $\langle z^{\tilde e}\rangle = \langle\exp(\tilde e\ln z)\rangle$ we are effectively summing $\tilde P(\tilde e)$ up to a cutoff around $\tilde e = 1/|\ln z|$. Because $\tilde P(\tilde e)\propto\tilde e^{1-\gamma}$ for large $\tilde e$, this sum will have a singular contribution scaling as $|\ln z|^{-(2-\gamma)}$. Then from (6) this is also the leading singular term for $z\to 1$ in $\langle z^k\rangle$, and hence $P(k)$ must have a power law tail with the same exponent as $\tilde P(\tilde e)$, i.e. $P(k)\propto k^{1-\gamma}$ for large k.

Numerical results for the degree distribution $P(k)$ are shown in Fig. 1, for networks of $N = 1,000$ neurons and several different combinations of the number of patterns per neuron α and the scale-free exponent γ determining $P(e)$. The results are consistent overall with the predicted $k^{1-\gamma}$ tail, though for larger α this is more difficult to see as it the tail behaviour is manifest only beyond an "central limit theorem" region for moderate k. This region arises because for large α each neuron can be connected to many patterns and its neuron-to-neuron degree is a sum over the degrees of these many patterns.

In the two graphs for $\gamma = 3.5$ in Fig. 1 we also show the theoretical prediction for $P(k)$ that is obtained by series expansion of (6), and find good agreement. For the two values of γ below 3, finite size effects are stronger: this is because here the power law tail $P(k)\propto k^{1-\gamma}$ means that the mean neuron-to-neuron degree is divergent, i.e. in practice grows with N. We do not show values of $\gamma < 2$, where $P(k)$ develops a non-normalizable $k^{1-\gamma}$ tail. The probability for any finite degree k must then go to zero for $N\to\infty$ so that $\mathcal{G}$ can no longer be regarded as sparse.

We also studied the degree of clustering of the scale-free neural networks $\mathcal{G}$ generated by Hebbian interactions. For each neuron i having $k_i\geq 2$ neighbors (in $\mathcal{G}$), the clustering coefficient $c_i = n_i/[k_i(k_i-1)/2]$ is the ratio of the number of actual edges among the neighbors, n_i, to the possible maximum $k_i(k_i-1)/2$. One can then look at the distribution of the c_i. More specifically we sort the c_i in order of increasing size and plot them against their normalized rank among the nodes of degree ≥ 2. This is essentially a cumulative histogram but with the x and y-axes swapped. Fig. 2 shows the results for the same combinations of γ and α as in Fig. 1. As a baseline we also show for $\gamma = 3.5$ the results for Erdös-Rényi graphs with the same average degree $\bar k = \alpha\langle e(e-1)\rangle$. We notice that compared to this baseline the local clustering coefficients exhibit a significantly wider spread, showing that the neural networks $\mathcal{G}$ are strongly heterogeneous. The clustering coefficients are also significantly larger than for Erdös-Rényi graphs; notice in particular that there is a finite fraction of neurons with the maximal clustering coefficient $c_i = 1$. This value will arise when, for example, a neuron i only occurs in a single pattern, as this pattern then automatically creates interactions between all of i's interaction partners. More generally, the high degree of clustering in $\mathcal{G}$ is related to the transitivity of the Hebbian coupling, which also tends to imply a degree of modularity [3]. In fact, modularity stems from groups of nodes sharing, in the bipartite network, connections with the same set of patterns; thus, modularity is larger for small values of α.

2. Accuracy of belief propagation on factor graphs

As is evident, one of the advantages of our present method is that we can easily investigate the processing capabilities of scale-free and highly clustered neural networks $\mathcal{G}$ such as the one in Fig. 1 (main text), i.e. by going via the associated bipartite network $\mathcal{B}$ of neurons and patterns as shown first in [4]. This implies that we are able to obtain the free energy (and related computational capabilities) for a wide class of networks, namely all those built via the Hebb rule. Above we focussed on the case where only the distribution $P(e)$ of the pattern degrees is constrained. However, our analysis in the main text applies also to the more general scenario where both $P(e)$ and the neuron(-to-pattern) degree distribution $P(d)$ are constrained.

We argued in the main text that our cavity or equivalently belief propagation analysis of information processing on such networks will become exact in the large N limit, and give further justification for this claim here. By construction, the factor graph ensembles we are considering have a prescribed degree profile, i.e. they belong to the well-known family of degree-constrained factor graphs [5] $\mathbb{D}_{N,P}(P(d), P(e))$, with N nodes, P patterns and prescribed degree profiles $P(d)$ and $P(e)$ for the neurons and patterns, respectively. This includes as a special case random regular factor graphs where the pattern degrees are homogeneous, i.e. fixed to some $\bar{e}$; correspondingly $P(\bar{e} = 1$ and $P(e) = 0$ for $e \neq \bar{e}$. Broadening $P(e)$ from this baseline makes the pattern degrees heterogeneous, with power-law distributions $P(e) \propto e^{-\gamma}$ representing significant, scale-free heterogeneity.

The key argument for our exactness of the cavity analysis for $N \to \infty$ is that such degree-constrained factor graphs become locally tree-like for $N \to \infty$. This has been shown rigorously [6], in the sense that $B_{i,r}(\mathbb{D})$, the neighborhood of finite radius r centered on the node i, converges for $N \to \infty$ in distribution to the neighborhood of a random tree graph $B_{i,r}(\mathbb{T})$, for each uniformly chosen random node i. This means that the number of finite-sized loops is thermodynamically irrelevant.

A complementary argument for the locally tree-like structure shows that the length of typical loops diverges for $N \to \infty$. In this respect it has been demonstrated [6, 7] that the typical distance from a given node where loops start playing a role is $O(\ln N)$. More specifically, let i be a uniformly chosen random node from a graph in $\mathbb{D}$, and let l_i be the length of the shortest loop through i, then with high probability,

$$l_i = \frac{\ln N}{\ln C}(1 + o(1)). \tag{7}$$

where $o(1)$ indicates terms going to zero for $N \to \infty$. The constant in the denominator is $C = B_d B_e$, the product of the branching ratios introducted in the main text. The result above applies for $C > 1$; for $C < 1$ loops are irrelevant anyway as the network will not percolate and consist of many disconnected tree-like pieces.

From (7) it is a standard argument [6] that Belief Propagation becomes exact for large N. Going beyond this, the leading corrections to the free energy – which are small, of $O(1/N)$ – that arise from residual loops have also been computed for a wide range of sparse graphs [8].

3. Asymmetric pattern distributions

In this section we sketch some details of our analysis for the case of asymmetric patterns. As explained in the main text, one introduces a degree of asymmetry $a \in [-1, +1]$ in the pattern distribution by taking for the nonzero pattern entries $P(\xi_i^\mu = \pm 1) = (1 \pm a)/2$.

Evaluating the ξ-average $\langle \Xi(\ldots) \rangle$ in eq. (8) of the main text, the condition for a transition to nonzero field means then becomes

$$1 = a^2 B_d \sum_e P(e) \frac{e(e-1)}{\langle e \rangle} \frac{\left\langle \sinh^2(z) \cosh^{e-2}(z) \right\rangle_z}{\left\langle \cosh^e(z) \right\rangle_z} \tag{8}$$

At zero temperature one argues as in the case of symmetric patterns that the averages over z in the numerator and denominator become dominated by large values of z so that their ratio approaches unity. The bifurcation then occurs when $B_d B_e = a^{-2}$; when a tends to zero the transition point goes to infinity and we retrieve the symmetric case. Beyond the bifurcation, the cavity field distributions become asymmetric (see Fig. 5 in the main text) and this leads to a non-zero global magnetization as is typical of ferromagnetic systems. One needs to bear in mind, however, that a bifurcation to growing field variances at zero means may compete with the bifurcation towards nonzero means and will give the physical transition if it occurs first on increasing β.

We evaluate numerically for both of the bifurcations at which temperature β_c^{-1} they occur. In both cases these bifurcation temperatures increase with α. To understand the asymptotic behaviour for large α one can then resort to a low-β expansion, namely $\left\langle \sinh^2(z) \cosh^{e-2}(z) \right\rangle \approx \left\langle z^2 \right\rangle = \beta$, $\langle \cosh^e(z) \rangle \approx 1$. This gives for the growing mean bifurcation condition (8) $1 \approx B_d B_e \beta a^2$ while for the growing variance bifurcation one gets $1 \approx B_d B_e \beta^2$ (see (12) in the main text). For Poisson graphs one has $B_d B_e = \alpha c^2$, giving the transition lines $\beta_{c,1}^{-1}(\alpha) \approx c^2 a^2 \alpha$ and $\beta_{c,2}^{-1}(\alpha) \approx c\sqrt{\alpha}$ for large α. In the presence of a nonzero pattern bias a these cross at $\alpha = 1/(ca^2)$, with the bifurcation to growing means occurring first for larger α. The existence of this crossing is confirmed by numerical evaluation of eq.(12) in the main text and (8) for finite α; the results are shown in Fig. 6 in the main text.

[1] S. Boccaletti, V. Latora, Y. Moreno, M. Chaves, *Phys. Rep.* **424**:4, 175, (2006).

[2] M. E. J. Newman, et al., *Phys. Rev. E*, **64(2)**:026118, (2001).

[3] E. Agliari, A. Barra, *A Hebbian approach to complex network generation*, Europhysics Letters **94**, 184201 (2013).

[4] A. Barra, F. Guerra, *J. Math. Phys.* **49**:125217, (2008).

[5] *Random graphs*, B.Bollobas (1998).

[6] *Information, Physics and Computation*, M.Mezard, A.Montanari /2009).

[7] *Models of random regular graphs*, N.C. Wormald-In Survey in combinatorics, Cambridge University Press (1997).

[8] U.Ferrari, C.Lucibello, F.Morone, G.Parisi, F.Ricci-Tersenghi, T.Rizzo- Physical Review B **88**, 184201 (2013).